

Information and communication in polygon theories.

Serge Massar* and Manas K. Patra*

March 13, 2014

Abstract

Generalized probabilistic theories (GPT) provide a framework in which one can formulate physical theories that includes classical and quantum theories, but also many other alternative theories. In order to compare different GPTs, we advocate an approach in which one views a state in a GPT as a resource, and quantifies the cost of interconverting between different such resources. We illustrate this approach on polygon theories (Janotta et al. New J. Phys 13, 063024, 2011) that interpolate (as the number n of edges of the polygon increases) between a classical trit (when $n = 3$) and a real quantum bit (when $n = \infty$). Our main results are that simulating the transmission of a single n -gon state requires more than one qubit, or more than $\log(\log(n))$ bits, and that n -gon states with n odd cannot be simulated by n' -gon states with n' even (for all n, n'). These results are obtained by showing that the classical capacity of a single n -gon state with n even is 1 bit, whereas it is larger than 1 bit when n is odd; by showing that transmitting a single n -gon state with n even violates information causality; and by showing studying the communication complexity cost of the nondeterministic not equal function using n -gon states.

1 Introduction.

The formalism of Generalised Probabilistic Theories (GPT) introduced several decades ago [1, 2, 3, 4, 5] provides a framework for studying generalisations of classical and quantum theory. The development of quantum information theory [6] motivated a renewed interest in GPT [7, 8, 9, 10, 11], with the aim of understanding from the point of view of information processing, what makes quantum theory special [12, 13]. Some phenomena considered as uniquely quantum, such as no-cloning and no-broadcasting, are generic features of GPT, see e.g. [8, 14, 15]. Considerable attention has also focused on

*Laboratoire d'Information Quantique, CP225, Department of Physics, Université libre de Bruxelles (ULB), Av. F. D. Roosevelt 50, B-1050 Bruxelles, Belgium. {smassar,manas.kumar.patra}@ulb.ac.be

theories more non-local than quantum [16, 17, 18, 19, 20], as no-signaling theories also have typically quantum properties such as intrinsic randomness, secret key generation, no-cloning, see e.g. [21, 22, 23]. On the other hand certain quantum features, such as continuity of operations [8], entanglement swapping and teleportation [24, 25], non triviality of communication complexity [17, 20], uniqueness of entropy [26, 27] and data processing inequalities [26] do not hold in many of these theories.

The study of quantum communication and entanglement has benefited from a quantitative approach in which one quantifies the cost of interconverting between different resources. For instance dense coding [28] and teleportation [29] show how quantum communication, entanglement, and classical communication can be interconverted. The amount of classical communication required to simulate the communication of a single qubit and the non-local correlations produced by a singlet have been established [30, 31, 32, 33, 35, 36, 34, 37], with lower bounds coming from communication complexity, see the review [38] for the case of many qubits, and [36] for the case of a single qubit. These interconversions show how much more powerful one resource is than another, and under what conditions are two resources equivalent.

We advocate here that a similar approach will be extremely fruitful in the study of GPT. In this approach one views the communication of a GPT state as a resource, and one wishes to quantify how much this resource is worth. What is the cost of replacing the communication of one GPT state by classical bits, by qubits, by states of another GPT? This question is not altogether new, but has been considered essentially only in the context of interconversions between non local correlations [18, 19, 39, 40, 41]. However this approach can be applied to all GPTs, and to simpler problems such as one way communication of a GPT state. Prior works in this direction include [42] in which it was shown that one way communication using a GPT based on the completely positive cone requires exponential classical (and conjectured quantum) communication to simulate, and [43] in which one way communication of “hypercube bits” is discussed.

Quantifying the cost of interconversion between GPTs is realised through two complementary techniques. First one constructs explicit protocols describing how GPT1 can be simulated by GPT2. Second one proves lower bounds on how many resources of GPT1 are required to simulate GPT2. The second kind of result will generally be obtained by exhibiting an information processing task that requires many resources using GPT1, but can be done cheaply using GPT2. Such tasks are very interesting, because they can discriminate between theories. They could therefore be used as physical/information theoretic arguments for selecting theories from the large space of GPTs.

Here we illustrate this approach in the context of polygon theories [44], so named because their state space is given by a regular polygon. We refer to them as n -gon theories, where n denotes the number of vertices of the polygon. This family of GPTs interpolates between a classical trit (when $n = 3$) and a qubit (when $n = \infty$). The non locality of these theories was studied in [44], where it was shown that there are profound differences between polygon theories

with odd and even number of vertices: the odd theories approach Tsirelson's bound from below, while even theories approach it from above. These results were however given without proof of tightness, and required a choice of tensor product.

Here we consider one way communication with polygon theories. This is a simpler context than [44], as the state space is smaller, and it does not require the introduction of a tensor product structure. For this reason one can hope to obtain more detailed results. Our main results are:

- For all n , simulating the communication of an n -gon state cannot be done with a single qubit.
- If the two parties do not have any shared randomness, we exhibit a protocol to simulate the communication of an n -gon state by sending $\log n$ bits. We also prove a lower bound of $\log(\log n)$ bits for this classical simulation.
- The n -gon theories with n odd cannot be simulated by the theories with n even.

These results are obtained by studying the classical capacity of n -gon states, the non deterministic communication complexity of the Not Equal function, and the communication complexity of the index function (also known as random access coding) with n -gon states. In the conclusion we discuss the open questions.

2 Polygon theories

We recall here the polygon theories introduced in [44] to which we refer for further details. The set of normalised states of n -gon theory has n extremal states:

$$\omega_i = \begin{pmatrix} r_n \cos \frac{2\pi i}{n} \\ r_n \sin \frac{2\pi i}{n} \\ 1 \end{pmatrix} \in \mathbb{R}^3 \quad , \quad r_n = \sqrt{\sec(\pi/n)} \quad , \quad 0 \leq i < n \quad (1)$$

The space of normalised states is the convex hull of $\{\omega_k\}$. It is a regular n -gon, thereby giving their name to these theories. The space of unnormalised states is the cone C_n generated by the $\{\omega_k\}$.

The space of effects is the dual C_n^* of the cone C_n . For even n the dual cone C_n^* is generated by the extremal effects

$$e_j = \frac{1}{2} \begin{pmatrix} r_n \cos \frac{(2j-1)\pi}{n} \\ r_n \sin \frac{(2j-1)\pi}{n} \\ 1 \end{pmatrix} . \quad (2)$$

In the even case the cone C_n is weakly self dual: the cone C_n^* equals the cone C_n rotated by the angle π/n .

For odd n the cone $C_n^* = C_n$ is generated by the extremal effects

$$e_j = \frac{1}{1+r_n^2} \begin{pmatrix} r_n \cos(2\pi j/n) \\ r_n \sin(2\pi j/n) \\ 1 \end{pmatrix} \quad (3)$$

For odd n the cone C_n is strongly self dual: $C_n^* = C_n$.

The unit effect is

$$u = (0, 0, 1) \ .$$

The normalised states have unit scalar product with the unit: $\langle \omega_k, u \rangle = 1$.

A measurement

$$M = \{f_k \in C_n^* : \sum_k f_k = u\}$$

is a set of effects (of elements of the dual cone) that sum to the unit effect. The probability of outcome k given normalised state ω and measurement M is

$$P(k|\omega) = \langle f_k, \omega \rangle \ .$$

In general a measurement can have an arbitrarily large number of effects. However in [42] it was shown that by refining a measurement, and decomposing a measurement into a convex combination of other measurements, one can restrict to measurements with at most 3 effects all proportional to the extremal effects. (Here 3 is the dimension of the space in which the states and effects are defined). That is we can restrict to measurements of the form:

$$M = \{\lambda_k e_k, k = 1, 2, 3 : 0 \leq \lambda_k, \sum_{k=1}^3 \lambda_k e_k = u\} \ . \quad (4)$$

In appendix A we present in more detail the structure of measurements with 3 extremal effects.

For n even the complement $u - e_i$ of any extremal effect e_i is an extremal effect. So in this case the measurements $M_i = \{e_i, u - e_i = e_{i+n/2}\}$ constitute 2-outcome measurements. (For odd n there do not exist 2 outcome measurements both of whose effects are extremal).

Note that the extremal effects eqs. (2, 3) are normalised such that $0 \leq \langle e_j, \omega_i \rangle \leq 1$. These inequalities are saturated in the following cases:

$$\langle e_j, \omega_i \rangle = 0 \quad \text{when} \quad i = j + n/2, \ i = j + n/2 - 1 \quad (n \text{ even}) \quad (5)$$

$$i = j + (n+1)/2, \ i = j + (n-1)/2 \quad (n \text{ odd}) \quad (6)$$

$$\langle e_j, \omega_i \rangle = 1 \quad \text{when} \quad i = j, \ i = j - 1 \quad (n \text{ even}) \quad (7)$$

$$i = j \quad (n \text{ odd}) \quad (8)$$

There are important differences between polygon theories with odd and even number of states. Some of these differences are discussed in [44] in the context of non-locality. They will also appear clearly in what follows.

3 Classical information capacity of polygon theories.

3.1 Statement of the result

We study the classical capacity of the n -gon theories. Alice receives symbols x drawn from a probability distribution p_x . Upon receiving symbol x , she sends state $\omega(x)$ to Bob, where $\omega(x)$ is a state of n -gon theory. Bob carries out a measurement, obtaining outcome y . The capacity of the channel is the maximum, over all probability distributions p_x , all encodings $\omega(x)$, all measurements, of the mutual information $I(X;Y)$. More precisely $I(X;Y)$ is, in the asymptotic limit, the number of bits Alice can send to Bob using many copies of the channel, using block coding, but no entanglement between systems.

This general setting can be simplified by noting that Bob's measurement can be taken from the set of canonical measurements eq. (4) (this follows from the data processing inequality, and the fact that the mutual information is convex in the probabilities $p_{y|x}$). These measurements have at most 3 outcomes, and hence the capacity of n -gon theories is at most $\log 3$ bits. Here we prove stronger results:

Theorem 1. *When n is even, the classical capacity of n -gon theories is exactly 1 bit.*

Theorem 2. *When n is odd, the classical capacity of n -gon theories is strictly larger than 1 bit, is equal to $\log 3$ bits when $n = 3$, and tends towards 1 bit when $n \rightarrow \infty$.*

To prove these results, we first exhibit communication protocols that satisfy the capacities stated in the theorems.

Protocol 1. (n even). *Alice has two inputs, $x = 0, 1$, that occur with equal probability $p_0 = p_1 = 1/2$. On input $x = 0$ she prepares state ω_0 , on input $x = 1$ she prepares state $\omega_{n/2}$. Bob carries out the measurement $M = \{e_0, e_{n/2}\}$.*

One easily checks using eqs. (8) that in this case the capacity is 1 bit.

Protocol 2. (n odd). *Alice has three inputs, $x = 0, 1, 2$, that occur with probability $p_0 = 1/2, p_1 = p_2 = 1/4$. On input $x = 0$ she prepares state ω_0 , on input $x = 1$ she prepares state $\omega_{(n-1)/2}$, on input $x = 2$ she prepares state $\omega_{(n+1)/2}$. Bob carries out the measurement $M = \{\frac{1}{\sqrt{n}}e_0, \frac{1}{2}e_{(n-1)/2}, \frac{1}{2}e_{(n+1)/2}\}$.*

Using eqs. (8) one checks that in this case the capacity is strictly larger than 1 bit. When $n = 3$ the capacity of this protocol is $\log 3$, and it decreases monotonically towards 1 bit as n tends to ∞ .

We now turn to proving the converses.

3.2 Upper bounds on the classical capacities.

As noted above, without loss of generality we can take Bob's measurement $M = \{\lambda_y e_y, y = 1, 2, 3\}$ to have 3 extremal effects, see eq. (4). The normalisation condition $\sum_y \lambda_y e_y = u$ implies that $\sum_y \lambda_y = 2$ (n even) and $\sum_y \lambda_y = 1 + r_n^2$ (n odd). For future notation, we denote $\sum_y \lambda_y = c_n$, where $2 \leq c_n \leq 3$. The probability of Bob obtaining outcome y given state $\omega(x)$ is $P(y|x) = \lambda_y \langle e_y, \omega(x) \rangle \leq \lambda_y$.

The mutual information $I(Y; X)$ is concave in the conditional probabilities $P(y|x)$ (i.e. taking a convex combination of two channels can only decrease the capacity). Hence the capacity is maximal at the extreme points of the set of conditional probabilities $P(y|x)$. If we enlarge the space of possible channels, we can only increase the capacity. In particular if we impose only the following conditions:

$$P(y|x) \geq 0 \quad (9)$$

$$P(y|x) \leq \lambda_y \quad (10)$$

$$\sum_y P(y|x) = 1 \quad (11)$$

$$\sum_y \lambda_y = c_n \quad (12)$$

with arbitrary alphabet size and number of measurement outcomes equal to three $y = 1, 2, 3$, then we include all channels realised by polygon theories (as well as some other channels). Hence the capacity of the channels defined by eqs. (9, 10, 11, 12) is at least as large as the capacity of the polygon theories.

Concavity of $I(X; Y)$ implies that the capacity of the channels defined by eqs. (9, 10, 11, 12) describe (for fixed alphabet size $|X|$) a polytope whose vertices are the extreme points. Concavity of $I(X; Y)$ implies that the capacity of the corresponding channels will be maximum at a vertex of this polytope. It is therefore sufficient to determine the capacities of the channels defined by the vertices of this polytope.

Lemma 1. *The vertices of the polytope defined by eqs. (9,11,10,12) have the following properties: Either at least one of the $\lambda_y = 0$; or when $2 < c_n \leq 3$, the vertices have (up to a permutation of the y 's) the form $\lambda_1 = c_n - 2$, $\lambda_2 = \lambda_3 = 1$, and all inputs x give rise to one of the following four output distributions:*

$$\begin{aligned} P(1|x_1) = 0 \quad , \quad P(2|x_1) = 0 \quad , \quad P(3|x_1) = 1 \\ P(1|x_2) = 0 \quad , \quad P(2|x_2) = 1 \quad , \quad P(3|x_2) = 0 \\ P(1|x_3) = c_n - 2 \quad , \quad P(2|x_3) = 0 \quad , \quad P(3|x_3) = 3 - c_n \\ P(1|x_4) = c_n - 2 \quad , \quad P(2|x_4) = 3 - c_n \quad , \quad P(3|x_4) = 0 . \end{aligned} \quad (13)$$

Note that Lemma 1 immediately implies the upper bounds stated in Theorems 1 and 2, since for $c_n = 2$ it implies that the capacity of the channel is at most 1 bit (since at least one outcome never occurs), and that for $3 \geq c_n > 2$,

the vertices either have capacity less or equal to 1 bit, or tend towards a channel with capacity 1 bit as $c_n \rightarrow 2$.

The proof of lemma 1 is given in appendix B. An alternative proof that the capacity of the n -gon theories with n even is bounded by 1 bit is given in appendix C.

3.3 Optimal coding for Alice

We also note that it is possible to simplify Alice's coding.

Lemma 2. *The maximal information capacity of polygon theories is obtained when Alice's alphabet has size 3, and each of the inputs is extremal $\omega(x) = \omega_{i(x)}$, $x = 1, 2, 3$.*

The proof of lemma 2 is given in appendix D.

4 Random access coding and information causality

We consider the task in which Alice receives uniformly random inputs of m bits, $x_1 x_2 \dots x_m$ and Bob receives as input a random index $j \in \{1, 2, \dots, m\}$. Bob's aim is to output a bit y that coincides with x_j . Depending on the context, this task goes under the name random access coding (RAC) [45], communication complexity of the index function [43], or information causality (IC)[46] .

Here, following [46], we shall measure the success of the protocol by the average information

$$\bar{I} = \sum_j p_j I_j = \sum_j p_j I(X : Y | j) \quad (14)$$

where $I_j = I(X : Y | j)$ is the conditional mutual information between Alice and Bob given that his input is j and $\{p_j\}$ is the probability distribution over Bob's input. Note that Fano's inequality [47] implies a relation between I_j and the probability P_j^{succ} of Bob successfully decoding Alice's j th input through $H(P_j) \geq 1 - I_j$.

As shown in [46], if Alice sends Bob c classical bits, or c qubits, then $I \leq c$ (even if Alice and Bob have shared randomness or shared entanglement). It can also be shown [48] by adapting slightly the proof in [46] that if Alice and Bob do not have prior shared entanglement, and Alice sends Bob q quantum bits, then $I \leq q$. The essential property used in these proofs is that the data processing inequality is valid both for classical [47] and quantum information theories [6].

The idea that I should be less than the classical capacity of the channel between Alice and Bob is known as information causality. Information causality does not hold in all GPTs, for instance if Alice and Bob share correlations more non local than quantum, and Alice sends Bob classical information.

Here we consider information causality in the context of n -gon theories (with n even). Specifically we suppose that Alice and Bob have as prior resource shared randomness. We consider the case $m = 2$: Alice receives two bits as input and sends Bob a state $\omega(x_1, x_2)$. Bob receives as input an index $j \in \{1, 2\}$. The aim is for Bob's output to maximize the quantity eq. (14).

Since the classical capacity of n -gon theories with n even is 1 bit (theorem 1) one would expect that $I = 1$. We show here that this intuition is wrong.

Theorem 3. *When n is even, sending a single state of n -gon theory, achieves $I > 1$.*

We conjecture that the maximum achievable value of I decreases monotonically from $I = 2$ when $n = 4$ to $I = 1$ when $n = \infty$ (because the limiting cases ($n = 4$ and $n = \infty$) are known and the explicit protocol exhibits this monotonicity).

Protocol 3. (n even). *Alice receives two bits x_0 and x_1 as input. She prepares the state $\omega_{x_0n/2+x_1+x_0x_1}$. Bob measures in the basis $M_1 = \{e_1, e_{n/2+1}\}$ and $M_0 = \{e_0, e_{n/2}\}$ on his inputs 0 and 1 respectively.*

Using this protocol, when $j = 0$, Bob learns the value of x_0 with certainty ($P(y = x_0 | j = 0) = 1$); while when $j = 1$ Bob's success probability is easily computed to be $P_n = 1 - \cos(2\pi/n)/2 > 1/2$. Therefore, $I > 1$.

5 Nondeterministic Not Equal Function

We recall a result from classical communication complexity (see [49] for a review of the field). Consider the NOT-EQUAL function, $F_{NE} : \{0, 1\}^k \times \{0, 1\}^k \rightarrow \{0, 1\}$ defined as

$$\begin{aligned} F_{NE}(x, y) &= 1 \text{ if } x \neq y \\ &= 0 \text{ if } x = y \end{aligned} \tag{15}$$

Suppose that Alice and Bob are given x and y , respectively, as inputs and their goal is to evaluate $F_{NE}(x, y)$ in the following weak sense. Bob should output a bit b that is distributed so that if $F_{NE}(x, y) = 0$, then $Pr[b = 1] = 0$; whereas if $F_{NE}(x, y) = 1$, then $Pr[b = 1] > 0$. In addition, assume that Alice and Bob have no prior shared randomness. Such a protocol can be regarded as a nondeterministic protocol for the F_{NE} function. The following lower bound on the amount of classical communication required by Alice and Bob to achieve this is well known (see [49]):

Lemma 3. *Any nondeterministic classical protocol for computing the function F_{NE} requires at least $\log_2(k)$ bits of communication.*

On the other hand we recall the result [36]:

Lemma 4. *There exists a nondeterministic quantum protocol for computing the function F_{NE} that uses one qubit of communication.*

We now show how polygon theories can be used to compute nondeterministically the NE function.

Lemma 5. *By sending a single state of n -gon theory, one can achieve a nondeterministic protocol for computing the function $F_{NE} : \{0, 1, \dots, n-1\} \times \{0, 1, \dots, n-1\} \rightarrow \{0, 1\}$.*

This is achieved by the following protocol.

Protocol 4. *Upon receiving input $x \in \{0, 1, \dots, n-1\}$, Alice sends Bob state ω_x . Upon receiving input $y \in \{0, 1, \dots, n-1\}$, Bob carries out measurement $M_y = \{e_y, e_{y+n/2}\}$ (n even) or $M_y = \{\{\frac{1}{r_n^2}e_y, \frac{1}{2}e_{y+(n-1)/2}, \frac{1}{2}e_{y+(n+1)/2}\}$ (n odd). Bob outputs 0 if he obtains outcome e_y , and outputs 1 if he obtains one of the other outcomes.*

One easily checks, using eq. (8) that this is a nondeterministic protocol for the NE function.

This implies the following lower bound on the classical cost of simulating polygon theories:

Theorem 4. *If Alice and Bob do not have a priori shared randomness, then simulating the communication of a single state of n -gon theory requires $\log_2(\log_2 n)$ bits of classical communication.*

On the other hand we have the following result:

Theorem 5. *If Alice and Bob do not have prior shared randomness, they can simulate the communication of a single state of n -gon theory using $\log_2 n$ bits of classical communication.*

which is achieved using the following simple protocol:

Protocol 5. *Suppose we wish to simulate Alice sending to Bob state ω , and Bob carrying out measurement $M = \{f_k\}$ on the state, such that result k is obtained with probability $P(k|\omega) = \langle f_k, \omega \rangle$. This can be simulated classically as follows. Alice decomposes $\omega = \sum_{i=0}^{n-1} p_i \omega_i$ into extremal states. She chooses (using local randomness) index $i \in \{0, \dots, n-1\}$ from the probability distribution p_i . She sends index i to Bob. Bob outputs k with probability $P(k|i) = \langle f_k, e_i \rangle$.*

6 Conclusion

In the present work we view generalised probabilistic theories as resources, and inquire how much of one resource is needed to simulate the other. This approach to studying GPT is particularly interesting because it immediately raises a large number of precise, quantitative, questions. To illustrate this approach we considered the specific case of polygon theories, and more particularly the situation where one n -gon state is sent from Alice to Bob. We have obtained a number of results for this family of theories, but many questions remain open.

Some of the main open questions in the context of n -gon theories are:

1. The transmission of one n -gon state cannot be simulated by the transmission of one qubit. How many qubits are in fact needed?
2. We have shown that in the absence of shared randomness, the cost of classically simulating the transmission of an n -gon state requires at least $\log_2(\log_2 n)$ bits. We have given a protocol that uses $\log n$ bits, and does not use any local randomness. What is the classical cost of simulating the transmission of an n -gon state with no randomness, with local randomness only, with shared randomness? Concerning the last question, it is not obvious whether one can transpose the protocols for the simulation of a single qubit [30, 31, 32, 33, 35, 36, 34, 37], since n -gon states cannot be simulated by a single qubit.
3. What is the cost of simulating an n -gon state by m -gon states? Our results on the classical capacities imply that a single n -gon states with n odd cannot be simulated by a single m -gon state with m even; and our results on the NOT EQUAL function imply that a single n -gon state cannot be simulated by a single m -gon state when $m < \log n$. However much tighter bounds may be possible.

A Extremal 3 effect measurements.

The following lemma provides the detailed structure of 3-outcome measurements all of whose effects are extremal.

Lemma 6. *In polygon theories, the measurements with three extremal effects are given by*

$$\begin{aligned}
M &= a(\lambda_1 e_{j_1}, \lambda_2 e_{j_2}, \lambda_3 e_{j_3}) \text{ where } 0 \leq \lambda_i \leq 1 \text{ and} \\
\lambda_1 &= 1 - \cot \frac{(j_1 - j_2)\pi}{n} \cot \frac{(j_3 - j_1)\pi}{n}, \\
\lambda_2 &= 1 - \cot \frac{(j_1 - j_2)\pi}{n} \cot \frac{(j_2 - j_3)\pi}{n}, \\
\lambda_3 &= 1 - \cot \frac{(j_2 - j_3)\pi}{n} \cot \frac{(j_3 - j_1)\pi}{n}, \\
a &= \begin{cases} 1 & (n \text{ even}) \\ (1 + r_n^2)/2 & (n \text{ odd}) \end{cases}
\end{aligned} \tag{16}$$

Further, for odd n there is constant $\delta_n > 0$ such that $\lambda_i \geq \delta_n$ and $\lim_{n \rightarrow \infty} \delta_n = 0$.

Proof. The coefficients λ_i and a follow after a tedious but straightforward calculation. The first part of the last statement follows from the fact that for odd n there do not exist measurements with 2 extremal effects. The limiting value is easily computed using the formulas in (16) and taking $j_1 = 0$, $j_3 = (n+1)/2$ and $j_2 = (n \pm 1)/4$, whichever is an integer. \square

B Proof of lemma 1.

Here we prove lemma 1 .

Proof. The number of variables in eqs. (9,11,10,12) is $3|X| + 3$. Taking into account the equalities eqs. (11,12), there are $2|X| + 2$ independent variables. A point $\{P(x|y), \lambda_y : x = 1, \dots, |X|, y = 1, 2, 3\}$ is a vertex if all inequalities and equalities in eqs. (9,11,10,12) are satisfied, and $2|X| + 2$ linearly independent inequalities eqs. (9,11) are saturated such that the values of all variables $P(x|y)$, λ_y are fixed.

Let us consider a specific input $x = x_0$. For this value of x_0 , there are 6 inequalities eqs. (9,11) that could be saturated. A sufficient number of them must be saturated that all values of $P(y|x_0)$ are fixed. We now enumerate all combinations of inequalities that can be saturated, such that they fix all the values of the $P(y|x_0)$. We give the inequalities that are saturated, and the implications for the other variables. For ease of notation, we give the inequalities that are saturated up to a permutation of the y 's.

2. Exactly 2 inequalities are saturated

- (a) $P(1|x_0) = P(2|x_0) = 0$ implies $P(3|x_0) = 1$ and $\lambda_3 > 1$.
- (b) $P(1|x_0) = 0, P(2|x_0) = \lambda_2$ implies $P(3|x_0) = 1 - \lambda_2, 0 < \lambda_2 < 1, \lambda_3 > 1 - \lambda_2$.
- (c) $P(1|x_0) = \lambda_1, P(2|x_0) = \lambda_2$ implies $P(3|x_0) = 1 - \lambda_1 - \lambda_2, 0 < \lambda_1 < 1, 0 < \lambda_2 < 1, \lambda_3 > 1 - \lambda_1 - \lambda_2$.

3. Exactly 3 inequalities are saturated

- (a) $P(1|x_0) = P(2|x_0) = 0, \lambda_3 = P(3|x_0) = 1$ implies $\lambda_1 > 0, \lambda_2 > 0, \lambda_1 + \lambda_2 = c_n - 1$.
- (b) $\lambda_1 = P(1|x_0) = P(2|x_0) = 0$ implies $P(3|x_0) = 1, \lambda_3 > 1$.
- (c) $\lambda_1 = P(1|x_0) = 0, P(2|x_0) = \lambda_2$ implies $P(3|x_0) = 1 - \lambda_2, 1 > \lambda_2 > 0, \lambda_3 > 1 - \lambda_2$.
- (d) $P(1|x_0) = 0, P(2|x_0) = \lambda_2, P(3|x_0) = \lambda_3$ implies $\lambda_1 > 0, \lambda_2 + \lambda_3 = 1$.

4. Exactly 4 inequalities are saturated

- (a) $\lambda_1 = \lambda_2 = P(1|x_0) = P(2|x_0) = 0$ implies $P(3|x_0) = 1, \lambda_3 = c_n$.
- (b) $\lambda_1 = P(1|x_0) = P(2|x_0) = 0, P(3|x_0) = \lambda_3$ implies $P(3|x_0) = 1, \lambda_3 = 1, \lambda_2 = c_n - 1$.

Now we turn to the variables λ_y . At a vertex it must be that the saturated inequalities fix the values of all the λ_y . We wish to find the vertices for which none of the λ_y 's vanish. Note that cases 3b,c and 4a,b fix at least one of the λ_y 's to zero, hence we cannot use these cases. Cases 2a,b,c do not fix any of the λ_y 's. Case 3a fixes $\lambda_3 = 1$. Case 3d fixes that $\lambda_2 + \lambda_3 = 1$.

Case 3d cannot by itself fix the values of the λ_y 's. Hence we must use at least once (say for variable x_1) case 3a, and we have $\lambda_3 = 1$, $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_1 + \lambda_2 = c_n - 1$.

We now wish to fix λ_1 and λ_2 (both different from zero). To this end we must use either cases 3a or 3d (for another value of x , say x_2).

When $c_n = 2$, one easily checks that using cases 3a,d to fix another value of λ implies $\lambda_3 = 1$, $\lambda_2 = 1$, $\lambda_1 = 0$ (or a permutation thereof), and therefore one of the λ_y 's is equal to zero.

When $c_n > 2$, the only way to fix all the λ_y 's without one of them vanishing is to use case 3a again to fix $\lambda_2 = 1$. Hence we have $\lambda_3 = 1$, $\lambda_2 = 1$, $\lambda_1 = c_n - 2$.

Finally there may be additional values of x (say x_3 and x_4) for which we can use case 2b (none of the other cases are compatible with these values of the λ_y 's).

We thus obtain the four probability distributions given in eq. (13). \square

C Direct proof of the upper bound in Theorem 1

In this section we provide a direct proof that the classical capacity of n -gon theories with n even is bounded by 1 bit. Recall that the probability matrix $P = P(y|x)$ of n -gon theories, n even, obeys eqs. (9,10,11,12) and that we can restrict the size of the input and output alphabets to three ($x, y = 1, 2, 3$).

Claim. Any matrix P satisfying eqs. (9,10,11,12) with $c_n = 2$ can be written as a convex combination of the following matrices.

$$\begin{aligned} P_1 &= \begin{pmatrix} u_1 & u_2 & u_3 \\ 1-u_1 & 1-u_2 & 1-u_3 \\ 0 & 0 & 0 \end{pmatrix}, \\ P_2 &= \begin{pmatrix} v_1 & v_2 & v_3 \\ 0 & 0 & 0 \\ 1-v_1 & 1-v_2 & 1-v_3 \end{pmatrix}, \\ P_3 &= \begin{pmatrix} 0 & 0 & 0 \\ w_1 & w_2 & w_3 \\ 1-w_1 & 1-w_2 & 1-w_3 \end{pmatrix}, \end{aligned} \quad (17)$$

where $0 \leq u_i, v_i, w_i \leq 1$, i.e. $P = q_1 P_1 + q_2 P_2 + q_3 P_3$ with $0 \leq q_1, q_2, q_3 \leq 1$ and $q_1 + q_2 + q_3 = 1$.

That is for $x = 1, 2, 3$ we can write

$$\begin{aligned} P(1|x) &= q_1 u_x + q_2 v_x, \\ P(2|x) &= q_1 (1 - u_x) + q_3 w_x, \\ P(3|x) &= q_2 (1 - v_x) + q_3 (1 - w_x). \end{aligned} \quad (18)$$

Proof. Fix the values of $\lambda_1, \lambda_2, \lambda_3$. We shall show that we can satisfy eqs. (18) with the choice

$$q_1 = 1 - \lambda_3, q_2 = 1 - \lambda_2 \text{ and } q_3 = 1 - \lambda_1 .$$

Note that since $P(1|x) + P(2|x) + P(3|x) = q_1 + q_2 + q_3 = 1$, there is at least one value of y such that $P(y|x) \leq q_y$. Without loss of generality we take $P(1|x) \leq q_1$.

Note that we only have to satisfy only two of the equations in (18) since the third is automatically satisfied by normalization. The following relations are necessary and sufficient for the first two equations in (18) to be satisfiable.

$$\begin{aligned} 0 &\leq P(1|x) - q_1 u_j \leq q_2 , \\ q_1(1 - u_j) &\leq P(2|x) \leq q_1(1 - u_j) + q_3 . \end{aligned} \quad (19)$$

Clearly these conditions are necessary. To prove sufficiency (assuming $P(1|x) \leq q_1$) if the relations in the first line are true for some choice of u_j then $0 \leq v_j = (p_{1j} - q_1 u_j)/q_2 \leq 1$. Similarly if the second line holds then $0 \leq w_j = (p_{2j} - q_1(1 - u_j))/q_3 \leq 1$.

Next, it is easily verified that the relations (19) are equivalent to

$$\max\{0, \frac{P(1|x)}{q_1} - \frac{q_2}{q_1}\} \leq u_j \leq \frac{P(1|x)}{q_1} , \quad (20)$$

$$1 - \frac{P(2|x)}{q_1} \leq u_j \leq \min\{1, 1 - \frac{P(2|x)}{q_1} + \frac{q_3}{q_1}\} . \quad (21)$$

Eqs. (20) and (21) are separately satisfiable. In order that they be simultaneously satisfiable it is necessary and sufficient that

$$\max\{0, \frac{P(1|x)}{q_1} - \frac{q_2}{q_1}\} \leq \min\{1, 1 - \frac{P(2|x)}{q_1} + \frac{q_3}{q_1}\} , \quad (22)$$

$$1 - \frac{P(2|x)}{q_1} \leq \frac{P(1|x)}{q_1} . \quad (23)$$

The first condition can be shown to be satisfiable by showing that both of the expressions on the left is \leq to both on the right. From eqs. (18) it follows that $P(2|x) \leq q_1 + q_3$, and hence that $0 \leq (q_1 + q_3 - P(2|x))/q_1$. Also $\frac{P(1|x)}{q_1} - \frac{q_2}{q_1} \leq 1$ due to our assumption that $P(1|x) \leq q_1$. Finally $\frac{P(1|x)}{q_1} - \frac{q_2}{q_1} \leq 1 - \frac{P(2|x)}{q_1} + \frac{q_3}{q_1}$ since this condition is simply $P(1|x) + P(2|x) \leq q_1 + q_2 + q_3 = 1$ which is always true. The second condition (23) is equivalent $q_1 \leq P(1|x) + P(2|x)$ which as noted above is also true due to the choice of q_i . \square

Since the mutual information $I(X : Y)$ is a concave function of the conditional probabilities $P(y|x)$ for fixed input distribution [47], and denoting by I_k mutual information corresponding to conditional probability matrix P_k , we have

$$I(X : Y) \leq q_1 I_1 + q_2 I_2 + q_3 I_3 \leq 1 .$$

The last inequality follows from the fact that the matrices P_i are essentially 2-dimensional.

D Proof of lemma 2.

Here we prove lemma 2 .

Proof. For simplicity of notation, we denote a channel by $C = \{\omega_x, p_x, e_y\}$, i.e. the collection of states, probabilities, and effects that are used.

The argument is related to that used in [42] to show that extremal measurements have at most 3 effects. To prove it, first note that if $\omega(x)$ is not extremal, we can decompose it into extremal states $\omega(x) = \sum_i p_{i|x} \omega_i$. We can then compare the two channels $C = \{\omega_x, p_x, e_y\}$ and $C' = \{\omega_i, p_{xi} = p_{i|x} p_x, e_y\}$. Using the chain rule for mutual information, we have $I(Y; XI) = I(Y; X) + I(Y; I|X)$ [47] where $I(Y; X)$ is the capacity of the channel C and $I(Y; XI)$ is the capacity of the channel C' . Hence the capacity of the channel C' is larger than the capacity of the channel C .

We can therefore suppose without loss of generality that Alice sends the extremal states ω_i with probabilities $p_i > 0$ (all strictly positive), where $i = 1, \dots, m$. We denote the corresponding channel $C = \{\omega_i, p_i, e_y\}$. We consider the case $m > 3$, and will show that there exists a channel with $m = 3$ with capacity at least as big as that of C .

Denote by $\omega = \sum_{i=1}^m p_i \omega_i$ the average state sent by Alice. By Caratheodory's theorem [50] ω can be written as a convex combination of at most three ω_i : $\omega = \sum_{j \in J} q_j \omega_j$ where $J \subset \{1, \dots, m\}$, $|J| \leq 3$. Let $q = \min_{j \in J} \frac{p_j}{q_j}$. We have $1 > q > 0$ (because all the $p_i > 0$, all the $q_j > 0$, and $|J| < m$) .

We can rewrite

$$\omega = q \left(\sum_{j \in J} q_j \omega_j \right) + (1 - q) \left(\sum_{i=1}^m \frac{p_i - q q_i}{1 - q} \omega_i \right)$$

where we set $q_i = 0$ when $i \notin J$. From the definition of q , it follows that both terms in parenthesis sum to ω , and that the coefficients $(p_i - q q_i)/(1 - q) \geq 0$ are positive, with at least one value of i such that $p_i - q q_i = 0$.

By recurrence we can write

$$\omega = \sum_k q_k \left(\sum_{j \in J_k} q_{j|k} \omega_j \right) \quad (24)$$

where $J_k \subset \{1, \dots, m\}$, $|J_k| \leq 3$ and

$$\sum_{j \in J_k} q_{j|k} \omega_j = \omega, \quad (25)$$

and $q_k \geq 0$, $\sum_k q_k = 1$, $q_{j|k} \geq 0$, $\sum_j q_{j|k} = 1$.

We can now compare the capacities of the original channel $C = \{\omega_i, p_i, e_y\}$ and the channel $C' = \{\omega_j, q_{jk} = q_k q_{j|k}, e_y\}$. In the second channel, Alice first chooses k with probability q_k , and then sends state ω_j with probability $q_{j|k}$, where we have that only three $q_{j|k}$ are non zero.

The overall probability distribution of channel C' can be written as $p_{yjk} = p_{y|jk}q_{j|k}q_k$. Eqs. (24, 25) imply that y is independent of k : $p_{y|k} = \sum_{j \in J_k} p_{y|jk}q_{j|k} = \sum_{j \in J_k} q_{j|k} \langle e_y, \omega_j \rangle = \langle e_y, \omega \rangle = p_y$. The chain rule for mutual information then implies that the capacity for channel C' is $I(Y; JK) = I(Y; K) + I(Y; J|K) = I(Y; J|K) = \sum_k q_k I(Y; J|K = k)$. where we have used that Y is independent of K . Hence there is at least one value of k for which $I(Y; J|K = k) \geq I(Y; JK)$. This value of k corresponds to a channel with an alphabet of size 3.

On the other hand we can write $I(Y; JK) = I(Y; J) + I(Y; K|Y) \geq I(Y; J)$ where we can identify $I(Y; J)$ as the capacity of channel C . Hence there is at least one value of k for which $I(Y; J|K = k) \geq I(Y; J)$, i.e. there is a channel with an alphabet of size 3 that has capacity greater or equal to the capacity of channel C . \square

References

- [1] G. Mackey, Mathematical Foundations of Quantum Mechanics, Benjamin (1963).
- [2] E. B. Davies and J. T. Lewis, An operational approach to quantum probability, Comm. Math. Phys. 17, 239–260 (1970).
- [3] C. M. Edwards, The operational approach to quantum probability I, Comm. Math. Phys. 17, 207–230 (1971).
- [4] D. J. Foulis and C. H. Randall, Empirical logic and tensor products, in H. Neumann, (ed.), Interpretations and Foundations of Quantum Theory, Bibliographisches Institut, Wissenschaftsverlag, Mannheim (1981).
- [5] G. Ludwig, An Axiomatic Basis of Quantum Mechanics 1, 2, Springer–Verlag (1985, 1987).
- [6] M. A. Nielsen and I. L. Chuang, Quantum computation and quantum information. Cambridge university press. (2010).
- [7] L. Hardy, Quantum theory from five reasonable axioms arXiv: quant-ph/0101012v4 (2001)
- [8] J. Barrett, Information processing in general probabilistic theories, Phys. Rev. A. 75 032304 (2007) (arXiv:quant-ph/0508211v3).
- [9] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Probabilistic theories with purification, Phys. Rev. A 81, 062348 (2010)
- [10] G. Chiribella, G. M. D’Ariano, and P. Perinotti. Informational derivation of quantum theory. Phys. Rev. A 84, 012311-012350 (2011)
- [11] L. Masanes and M. P. Müller, A derivation of quantum theory from physical requirements, New J. Phys. 13, 063001 (2011)

- [12] C. A. Fuchs, Quantum mechanics as quantum information (and only a little more), arXiv preprint quant-ph/0205039 (2002)
- [13] G. Brassard. Is information the key?. *Nature Physics* 1, pp. 2-4 (2005).
- [14] H. Barnum, J. Barrett, M. Leifer and A. Wilce, Cloning and broadcasting in generic probabilistic models (2006) (arXiv:quant-ph/061129).
- [15] H. Barnum, J. Barrett, M. Leifer and A. Wilce, Generalized No-Broadcasting Theorem, *Phys. Rev. Lett.* 99, 240501 (2007) (arXiv:0707.0620).
- [16] S. Popescu and D. Rohrlich (1994). Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3), 379-385.
- [17] W. van Dam, Implausible consequences of superstrong nonlocality. *Natural Computing*, 1-4 (2013) (arXiv:quant-ph/0501159)
- [18] Barrett, J., Linden, N., Massar, S., Pironio, S., Popescu, S., and Roberts, D. (2005). Nonlocal correlations as an information-theoretic resource. *Physical Review A*, 71(2), 022101.
- [19] Barrett, J., and Pironio, S. (2005). Popescu-Rohrlich correlations as a unit of nonlocality. *Physical review letters*, 95(14), 140401.
- [20] Brassard, G., Buhrman, H., Linden, N., Méthot, A. A., Tapp, A., and Unger, F. (2006). Limit on nonlocality in any world in which communication complexity is not trivial. *Physical Review Letters*, 96(25), 250401.
- [21] L. Masanes, A. Acin, N. Gisin, General properties of nonsignaling theories, *Phys. Rev. A* 73, 012112 (2006)
- [22] J. Barrett, L. Hardy, A. Kent, No signaling and quantum key distribution. *Phys. Rev. Lett.* 95, 010503 (2005)
- [23] A. Acin, N. Gisin, L. Masanes, From Bell's theorem to secure quantum key distribution, *Phys. Rev. Lett.* 97, 120405 (2006).
- [24] H. Barnum, J. Barrett, M. Leifer and A. Wilce, Teleportation in general probabilistic theories, In *Proceedings of Symposia in Applied Mathematics* (Vol. 71, pp. 25-48) (2012) (arXiv:0805.3553).
- [25] H. Barnum, P. Gaebler and A. Wilce, Ensemble Steering, Weak Self-Duality, and the Structure of Probabilistic Theories (2009) (arXiv:0912.5532)
- [26] H. Barnum, J. Barrett, L. Orloff Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, R. Wilke, Entropy and Information Causality in General Probabilistic Theories, *New J. Phys.* 14, 129401 (2012)

- [27] Short, A. J. and Wehner, S. (2010). Entropy in general physical theories. *New Journal of Physics*, 12(3), 033023.
- [28] C. H. Bennett and S. J. Wiesner, Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states, *Phys. Rev. Lett.* 69, 2881–2884 (1992)
- [29] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres and W. K. Wootters, *Phys. Rev. Lett.* 70, 1895–1899 (1993)
- [30] T. Maudlin, in *PSA 1992, Volume 1*, edited by D. Hull, M. Forbes, and K. Okruhlik (Philosophy of Science Association, East Lansing, 1992), pp. 404–417.
- [31] G. Brassard, R. Cleve, and A. Tapp, *Phys. Rev. Lett.* 83, 1874 (1999).
- [32] M. Steiner, *Phys. Lett. A* 270, 239 (2000).
- [33] J. A. Csirik, *Phys. Rev. A* 66, 014302 (2002).
- [34] D. Bacon and B. F. Toner, *Phys. Rev. Lett.* 90, 157904 (2003).
- [35] N. J. Cerf, N. Gisin, and S. Massar, *Phys. Rev. Lett.* 84, 2521 (2000).
- [36] S. Massar, D. Bacon, N. Cerf, and R. Cleve, *Phys. Rev. A* 63, 052305 (2001)
- [37] Toner, B. F., and Bacon, D., Communication Cost of Simulating Bell Correlations. *Physical Review Letters*, 91(18), 187904 (2003).
- [38] Buhrman, H., Cleve, R., Massar, S., and de Wolf, R. (2010). Nonlocality and communication complexity. *Reviews of modern physics*, 82(1), 665.
- [39] Jones, N. S., and Masanes, L. (2005). Interconversion of nonlocal correlations. *Physical Review A*, 72(5), 052312.
- [40] Brunner, N., and Skrzypczyk, P., Nonlocality distillation and postquantum theories with trivial communication complexity. *Phys. Rev. Lett.* 102, 160403 (2009)
- [41] M. Forster, S. Winkler, and S. Wolf (2009). Distilling nonlocality. *Physical review letters*, 102(12), 120401.
- [42] Fiorini, S., Massar, S., Patra, M. K., and Tiwary, H. R. (2013). Generalised probabilistic theories and conic extensions of polytopes. *arXiv preprint arXiv:1310.4125*.
- [43] N. Brunner, M. Kaplan, A. Leverrier, P. Skrzypczyk, Dimension of physical systems, information processing, and thermodynamics, *arXiv:1401.4488*
- [44] P. Janotta, C. Gogolin, J. Barrett and N. Brunner, Limits on nonlocal correlations from the structure on the local state space, *New J. Phys.* **13**, 063024 (2011).

- [45] A. Ambainis, A. Nayak, A. Ta-Shma, U. Vazirani, Journal of the ACM, 49(4) 496 (2002).
- [46] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter and Marek Żukowski, Information causality as a physical principle, Nature 461, 1101-1104 (2009)
- [47] Thomas M. Cover and Joy A. Thomas, Elements of Information Theory, Wiley-Interscience, 1991
- [48] M. Pawłowski and V. Scarani, Information Causality, arXiv:1112.1142
- [49] E. Kushilevitz and N. Nisan, Communication Complexity. Cambridge University Press, Cambridge, England, 1998.
- [50] A. Barvinok, A Course in Convexity, Graduate Studies in Mathematics, V. 54 (2002)